Patent
File 235.00440101

DIPEPTIDYLPEPTIDASES AND METHODS OF USE

RELATED APPLICATIONS

This application claims the benefit of the U.S. Provisional Application No. 60/246,827, filed November 8, 2000, which is incorporated by reference in its entirety.

GOVERNMENT FUNDING

The present invention was made with government support under Grant No. DE 09761, awarded by the National Institutes of Health. The Government may have certain rights in this invention.

BACKGROUND

Porphyromonas gingivalis (*P. gingivalis*), an oral anaerobic bacterium, has been implicated as a causative agent of adult type periodontitis. As an asaccharolytic organism, *P. gingivalis* is totally dependent on external sources of peptides that are necessary for its growth and proliferation. In order to fulfill such a fastidious nutritional requirement this bacterium evolved a complex system of proteolytic enzymes which are now recognized as important virulence factors in the development of periodontal disease (Travis et al., *J. Adv. Exp. Med. Biol.*, 477:455-65 (2000)). The best known and well characterized enzymes of this system are gingipains R and K, arginine and lysine specific, cysteine proteinases (Curtis et al., *J. Periodontal Res.*, 34:464-72 (1999)). Working in concert with the proteinases periodontain (Nelson et al., *J. Biol. Chem.*, 274:12245-51 (1999)), collagenases/gelatinases (Birkedal-Hansen et al., *J. Periodontal Res.*, 23:258-64 (1988); Lawson et al., *Infect. Immun.*, 60:1524-29 (1992); Kato et al., *J. Bacteriol.*, 174:3889-95 (1992), prtT (Otogoto et al., *Infect. Immun.*, 61:117-23 (1993)), and Tpr (Bourgeau et al., *Infect. Immun.*, 60:3186-92 (1992)) as well as host proteinases, this array of enzymes has the

potential to degrade proteins from both the periodontal ligamentum and surrounding tissues. Their concerted action leads to the formation of a large pool of oligopeptides, which can be further utilized by *P. gingivalis* and other oral bacteria. However, *P. gingivalis* cannot transport poly- and oligo- peptides into the cell, even though it has the ability to thrive on dipeptides as a sole source of carbon. This has led to an interest in studying a specialized group of *P. gingivalis* peptidases capable of hydrolyzing oligopeptides to di- and tripeptides, which can be subsequently metabolized by this periodontopathogen. The purification, characterization and cloning of prolyl tripeptidylpeptidase A (PtpA), an enzyme which liberates tripeptides from the N-terminal regions of substrates containing proline residues in the third position has been previously reported (Banbula et al., *J. Biol. Chem.*, 274:9246-52 (1999)). Dipeptidylpeptidase-IV (DPP-IV), an enzyme with similar specificity, but only dipeptylpeptidase activity, has also been cloned (Kiyama et al., *Biochim. Biophys. Acta*, 1396:39-46 (1998)), purified, and characterized (Kumagai et al., *Infect. Immun.*, 68:716-24 (2000); Banbula et al., *Infect. Immun.*, 68:1176-82 (2000)). Together with a recently described angiotensinogen-converting enzyme analogue (Awano et al., *FEBS Lett.*, 460:139-44 (1999)) all of these proteases can hydrolyze peptide bonds containing proline residues. In addition, the *P. gingivalis* genome contains three further putative coding sequences encoding proteinases homologous with dipeptidylpeptidase-IV, although their activities have not yet been identified (Banbula et al., *J. Biol. Chem.*, 274:9246-52 (1999)).

SUMMARY OF THE INVENTION

In one aspect, the present invention provides an isolated dipeptidylpeptidase, active analog, active fragment, or active modification thereof having amidolytic activity for cleavage of a peptide bond between the second and third amino acids from the N-terminal end of a target polypeptide, wherein the target polypeptide has an aliphatic or an aromatic residue as a substituent on the α-carbon atom of the second amino acid from the N-terminal end of the polypeptide. Preferably, the dipeptidylpeptidase is isolated from *Porphyromonas gingivalis*. Preferably, the dipeptidylpeptidase is a serine

protease. Preferably, the dipeptidylpeptidase includes an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, and SEQ ID NO:26. Preferably the dipeptidylpeptidase is encoded by a nucleic acid including a nucleotide sequence SEQ ID NO:1.

In another aspect, the present invention provides an isolated polypeptide including an amino acid sequence having a percentage amino acid identity greater than about 40% with SEQ ID NO:2.

In another aspect, the present invention provides an isolated nucleic acid including a coding sequence encoding a dipeptidylpeptidase, active analog, active fragment, or active modification thereof having amidolytic activity for cleavage of a peptide bond between the second and third amino acids from the N-terminal end of a target polypeptide, wherein the target polypeptide has an aliphatic or an aromatic residue as a substituent on the α-carbon atom of the second amino acid from the N-terminal end of the polypeptide. Preferably the nucleic acid includes a nucleotide sequence SEQ ID NO:1. Alternatively, the complement of the nucleic acid preferably hybridizes to SEQ ID NO:1 under hybridization conditions of 0.5 M phosphate buffer, pH 7.2, 7% SDS, 10 mM EDTA, at 68°C, followed by three for 20 minutes washes in 2x SSC, and 0.1% SDS, at 65°C, wherein at least about 20 nucleotides of the complement hybridize.

In another aspect, the present invention provides an isolated nucleic acid encoding a polypeptide, wherein the polypeptide includes an amino acid sequence having a percentage amino acid identity greater than about 40% with SEQ ID NO:2.

In another aspect, the present invention provides a method of identifying an inhibitor of a dipeptidylpeptidase, active analog, active fragment, or active modification thereof. The method includes identifying a compound that inhibits the amidolytic activity of the dipeptidylpeptidase by incubating the dipeptidylpeptidase with the compound under conditions that promote amidolytic activity of the dipeptidylpeptidase and determining if the amidolytic

activity of the dipeptidylpeptidase is inhibited relative to the amidolytic activity in the absence of the compound.

In another aspect, the present invention provides a method of reducing growth of a bacterium including inhibiting a dipeptidylpeptidase, active analog, active fragment, or active modification thereof, by contacting the dipeptidylpeptidase with an inhibitor of the dipeptidylpeptidase. Preferably the dipeptidylpeptidase is a serine protease.

In another aspect, the present invention provides a method for protecting an animal from a periodontal disease caused by *Porphyromonas gingivalis* including administering to the animal an inhibitor of dipeptidylpeptidase, wherein the disease is selected from the group consisting of gingivitis and periodontitis. Preferably the inhibitor is administered by a method selected from the group consisting of subgingival application and controlled release delivery.

In another aspect, the present invention provides an immunogenic composition including an isolated dipeptidylpeptidase, an antigenic analog, an antigenic fragment, or an antigenic modification thereof having amidolytic activity for cleavage of a peptide bond present in a target polypeptide, the peptide bond being located between the second and third amino acids from the N-terminal end of the target polypeptide, wherein the second amino acid from the N-terminal end has an aliphatic or an aromatic residue as a substituent on the α-carbon atom. Preferably the dipeptidylpeptidase is a serine protease. Preferably the second amino acid is selected from the group consisting of alanine, phenylalanine, isoleucine, and leucine. The immunogenic composition may optionally include an adjuvant.

In another aspect, the present invention provides a composition including an inhibitor of an isolated dipeptidylpeptidase and a pharmaceutically acceptable carrier.

Definitions

"Polypeptide" as used herein refers to a polymer of amino acids and does not refer to a specific length of a polymer of amino acids. Thus, for example, the terms peptide, oligopeptide, protein, and enzyme are included within the definition of polypeptide. This term also includes post-expression modifications

of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. A polypeptide can be produced by an organism, or produced using recombinant techniques, or chemically or enzymatically synthesized.

"Polynucleotide" and "nucleic acid" are used herein interchangeably and refer to a linear polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides, and include both double- and single-stranded DNA and RNA. A nucleic acid may include both coding and non-coding regions that can be obtained directly from a natural source (e.g., a microorganism), or can be prepared with the aid of recombinant or synthetic techniques. A nucleic acid may be equivalent to this nucleic acid or it can include, in addition, one or more other polynucleotides. For example, the nucleic acid of the invention can be a vector, such as an expression of a coding sequence.

"Peptidase," "proteinase," and "protease" all refer to enzymes that catalyze the hydrolysis of peptide bonds in a polypeptide. A "peptide bond" or "amide bond" is a covalent bond between the alpha-amino group of one amino acid and the carboxyl group of another amino acid. "Peptidase inhibitor," "proteinase inhibitor," "protease inhibitor," and "inhibitor" all refer to compounds that inhibit a peptidase that catalyzes the hydrolysis of peptide bonds in a polypeptide.

"Serine protease" refers to an enzyme that uses the hydroxy-functional side chain of serine as a nucleophile in a catalytic reaction.

"Amidolytic activity" refers to the ability of a polypeptide to catalyze the hydrolysis of at least one peptide bond in a polypeptide. The term "cleavage" can also be used to refer to the hydrolysis of a peptide bond in a polypeptide. A "dipeptidylpeptidase" is able to hydrolyze the peptide bond between the second and third amino acids from the N-terminal end of a target polypeptide including the general formula H-Xaa-Yaa-Xaa-, wherein Xaa is a natural or modified amino acid, and Yaa is an amino acid including an aliphatic or an aromatic residue as a substituent on the α-carbon atom. Preferred amino acids in the Yaa position include alanine, phenylalanine, isoleucine, and leucine.

A "target polypeptide" is a polypeptide that is the potential substrate of the amidolytic activity of a dipeptidylpeptidase. A "dipeptidylpeptidase" does not have to cleave all members of the target polypeptide. The term "natural amino acid" refers to the 20 amino acids typically produced by a cell. The term "modified amino acid" refers to, for instance, acetylation, hydroxylation, methylation, amidation, or the attachment of carbohydrate or lipid moieties, cofactors, and the like.

As used herein, the term "isolated" means that a polypeptide or a polynucleotide has been either removed from its natural environment, produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, the polypeptide or polynucleotide is purified, i.e., essentially free from any other polypeptides, polynucleotides, and associated cellular products or other impurities.

An active analog, active fragment, or active modification of a polypeptide of the invention is one that has amidolytic activity by hydrolysis of a peptide bond present in the target polypeptide as described herein. Active analogs, active fragments, and active modifications are described in greater detail herein.

An antigenic analog, antigenic fragment, or antigenic modification of a polypeptide of the invention is one that has amidolytic activity by hydrolysis of a peptide bond present in the target polypeptide as described herein. Antigenic analogs, antigenic fragments, and antigenic modifications are described in greater detail herein.

"Percentage amino acid identity" refers to a comparison of the amino acids of two polypeptides as described herein.

As used herein, "aliphatic residue" means an organic radical having carbon atoms linked in open chains.

As used herein, "aromatic residue" means an organic radical that includes an aromatic ring (e.g., an aromatic group, an alkaryl group, or an aralkyl group).

As used herein, the "P1" position of a polypeptide is the amino acid on the N-terminal end of the scissile bond that is being cleaved. For dipeptidylpeptidases that cleave the peptide bond between the second and third

amino acids from the N-terminal end of a target polypeptide, the P1 position is the second amino acid from the N-terminal end of the target polypeptide (i.e., the penultimate position).

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts a plot of the absorbance at 280 nm (Δ) and amidolytic activity against Ala-Phe-pNA (•) for the purification of *P. gingvivalis* dipeptidylpeptidase (DPP-7) from the acetone precipitate of the *P. gingivalis* cell extract. The straight solid lines indicate gradients in the eluting composition. Figure 1*(a)* illustrates the separation of DPP-7 on hydroxyapatite (100 ml) equilibrated with 20 mM potassium phosphate buffer, pH 7.0, and using a potassium phosphate gradient from 20 mM to 300 mM. Figure 1*(b)* illustrates the separation of DPP-7 obtained from the previous step on Phenyl-Sepharose HP (25 ml) equilibrated with 50 mM potassium phosphate, 1M ammonium sulfate, pH 7.0, at a flow rate of 30 ml/hour, and using an ammonium sulfate gradient from 0.4M to 0M. Figure 1*(c)* illustrates the separation of DPP-7 on a MonoS FPLC column using a sodium chloride gradient from 0M to 0.3M then from 0.3M to 1M.

Figure 2 is a depiction of the SDS-PAGE of fractions obtained during the purification of *P. gingivalis* DPP-7 with *Lane A* representing molecular mass markers (phosphorylase B, 97 kDa; bovine serum albumin, 68 kDa; ovalbumin, 43 kDa; carbonic anhydrase, 30 kDa; soybean trypsin inhibitor, 20 kDa; α-lactalbumin, 14 kDa); *Lane B* representing acetone precipitate from Triton X-100 extract of *P. gingivalis; Lane C* representing hydroxyapatite column eluate; *Lane D* representing Phenyl-Sepharose column eluate; and *Lane E* representing MonoS column eluate.

Figure 3 depicts a plot of the DPP-7 activity against Ala-Phe-pNA vs. pH. Enzyme activity was tested on Ala-Phe-pNA substrate in different buffers including: HEPES (•); PIPES (□); potassium phosphate (■); Tris (○); and MES (▲).

Figure 4 depicts the coding sequence (SEQ ID NO:1) encoding *P. gingivalis* DPP-7 (SEQ ID NO:2). Sequences obtained from the Edman

degradation of the trypsin fragmented DPP-7 polypeptide chain are underlined. The putative active site serine residue is marked by the black background.

Figure 5 is a listing of sequences comparing the C-terminal regions of the *P. gingivalis* DPP-7 (residues 664-695; SEQ ID NO:3) and *S. aureus V8* endopeptidase (residues 704-863; SEQ ID NO:4). Common residues are indicated by the single letter amino acid in the line between the two sequences. The "+" symbol in the line between the two sequences indicates similar residues.

Figure 6 depicts a multiple sequence alignment of *P. gingivalis* DPP-7 and its putative homologues. Sequences of DPP-7 related proteinases were obtained from the conceptual translation of the following ORFs retrieved from unfinished and finished genomes databases (available at www.tigr.org): S1- *Shewanella putrefaciens* gnl |TIGR_24|sputre 6401 (SEQ ID NO:5); S2- *Shewanella putrefaciens* gnl |TIGR_24|sputre 6410 (SEQ ID NO:6); X- *Xylella fastidiosa* gb|AE004008.1| (SEQ ID NO:7); P1- *Porphyromonas gingivalis* gnl |TIGR|*P. gingivalis*_CPG.con (SEQ ID NO:8); P2- *P. gingivalis* DPP-7 gnl |TIGR|*P. gingivalis*_CPG.con (SEQ ID NO:9). The sequences were subsequently aligned using the ClustalW multiple sequence alignment tool.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides isolated polypeptides, preferably isolated dipeptidylpeptidases, that have amidolytic activity by hydrolysis of a peptide bond present in a target polypeptide, wherein the bond is between the second and third amino acid from the N-terminus of the peptide. The dipeptidylpeptidase has amidolytic activity by hydrolysis of a peptide bond present in a target polypeptide including the fragment H-Xaa-Yaa-Xaa-, wherein Xaa is a natural or modified amino acid, Yaa is an amino acid including an aliphatic or an aromatic residue as a substituent on the α-carbon atom, and the peptide bond of the target polypeptide that is hydrolyzed is the bond between the second and third amino acids from the N-terminus of the peptide. In increasing order of preference,

isolated polypeptides can cleave a target polypeptide that is at least about 5 amino acids or at least about 400 Da, at least about 10 amino acids or at least about 750 Da, at least about 20 amino acids or at least about 1,500 Da, or at least about 30 amino acids or at least about 3,000 Da. Preferably, the dipeptidylpeptidases cleave peptides including a sequence of H-Xaa-Yaa-Xaa-, wherein Yaa is alanine, phenylalanine, isoleucine, or leucine. More preferably, the dipeptidylpeptidases cleave peptides including a sequence of SEQ ID NO:10, SEQ ID NO:11; SEQ ID NO:12; or SEQ ID NO:13 as shown in Table 3.

The polypeptides disclosed in the present application are preferably dipeptidylpeptidases. Preferably, the dipeptidylpeptidase is isolated from *Porphyromonas gingivalis*. Preferably, the dipeptidylpeptidase is a serine protease. Most preferably, the dipeptidylpeptidase is *P. gingivalis* dipeptidylpeptidase-7 (DPP-7). The polypeptides can be used as a source of antibodies for inhibiting the amidolytic activity and thereby possibly reducing periodontitis, loss of tooth attachment and periodontal pocket formation. Antibodies to dipeptidylpeptidases can also be used to identify and/or isolate additional dipeptidylpeptidases. Knowledge of dipeptidylpeptidases can also be used to make inhibitors of dipeptidylpeptidases and to make immunogenic compositions that could be used to elicit the production of antibodies to dipeptidylpeptidases and thereby possibly reduce gingivitis, periodontitis, loss of tooth attachment, and/or periodontal pocket formation.

Dipeptidylpeptidase-7, either alone or in a mixture with other dipeptidylpeptidases, can be used to generate a pool of dipeptides from polypeptides. Dipeptides may be preferably imported by cells. Thus, pools of didpeptides might be useful substrates for transport.

Dipeptidylpeptidase-7 (DPP-7) was purified from the membrane fraction of *Porphyromonas gingivalis*. This enzyme, preferably having an apparent molecular mass of about 76 kDa, has specificity for polypeptides having either an aliphatic or an aromatic residue as a substituent on the α-carbon atom of the second amino acid from the N-terminal end of the polypeptide. Although it belongs to the serine class of peptidases, it does not resemble other known dipeptidylpeptidases. Interestingly, the amino acid sequence around the putative

active site serine residue shows significant homology to the C-terminal region of the *Staphylococcus aureus* V-8 endopeptidase. In *P. gingivalis*, DPP-7 probably serves nutritional functions by providing dipeptides to this assaccharolytic bacterium.

Several studies indicate that the outer membrane of *P. gingivalis* contains a complex, proteolytic machinery which serves multiple physiological functions. The present application discloses the identification of a novel proteinase localized on the bacterial surface.

The purified enzyme migrated as a single band of about 76 kDa on SDS-PAGE and its amino-terminal sequence was located within the primary structure of the translated product of the *dpp-7* coding sequence. Apparently, the enzyme is truncated at the amino terminus (i.e., amino acid 24 of SEQ ID NO:2 is the first amino acid of the truncated form) due to the action of a lysine specific proteinase, most likely gingipain K. Taking into account that the N-terminus of DPP-7 contains membrane anchorage domains it is likely that the N-terminal truncation noted here occurred during the isolation procedure and may not represent its true membrane form.

The dipeptidylpeptidases of the present invention are preferably serine proteases that are inhibited by serine protease inhibitors. The dipeptidylpeptidases of the present invention are preferably inhibited by serine protease inhibitors including, for example, diisopropylfluorophosphate (DFP), 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (PEFABLOCK), and phenylmethanesulfonyl fluoride (PMSF). In addition, the dipeptidylpeptidases of the present invention are resistant to sulfhydryl group blocking reagents and chelating agents, which is also consistent with the enzyme being a serine protease. However, the *P. gingivalis* DPP-7 does not belong to any of the six previously described types of dipeptidylpeptidases (Barrett et al., *Handbook of Proteolytic Enzymes*, Academic Press, London (1998)). DPP-I is a member of a cysteine class of peptidases and possesses a broad specificity, but has an exclusion for basic amino acid and proline residues in the P1 site of the scissile peptide bond (McGuire et al., *Arch. Biochem. Biophys.*, 295:280-88 (1992)). DPP-VI is another representative of the cysteine proteinase family with

dipeptidylpeptidase activity towards broad spectrum of substrates (Vacheron et al., *Eur. J. Biochem.*, 100:189-96 (1979)). DPP-II, DPP-IV and DPP-V belong to the S9 family of the serine proteases (Barrett et al., *Handbook of Proteolytic Enzymes*, Academic Press, London (1998)). Both DPP-II and DPP-IV share similar specificity directed against Pro and Ala residues in the penultimate position whereas DPP-V is an enzyme secreted by *Aspergillus fumigatus* with a unique substrate specificity limited to X-Ala, His-Ser, and Ser-Tyr dipeptides (Beauvais et al., *J. Biol. Chem.*, 272:6238-44 (1997)). DPP-III is also classified as a serine peptidase, with its action being restricted to Arg residue in the P1 position (Ellis et al., *J. Biol. Chem.*, 242:4623-29 (1967)). In terms of biochemical features, DPP-7 resembles a dipeptidyl aminopeptidase (DAP-BII), which was isolated from *Pseudomonas sp.* strain WO24, but the coding sequence of that enzyme remains unknown and does not allow a sequence comparison of these proteins (Ogasawara et al., *J. Bacteriol.*, 178:6288-95 (1996)). Because *P. gingivalis* dipeptidylpeptidase does not exhibit any significant homology to any of the dipeptidylpeptidases described so far, this enzyme has been designated DPP-7.

Interestingly the *P. gingivalis* DPP-7 displays the consensus sequence characteristic for the catalytic site of the V-8 like proteases, a group of endopeptidases cleaving after glutamic or aspartic acid residues (Carmona et al., *Nucleic Acids Res.*, 15:6757 (1987)). This region of homology is specifically located only at the C-terminal region of both proteases and includes the putative active site serine residue. Interestingly, more coding sequences encoding putative, DPP-7 related proteases in *P. gingivalis*, *Xylella fastidiosa* and *Shewanella putrefaciens* were identified. Based on the enzymological and coding sequence data presented above, the *P. gingivalis* DPP-7 does not belong to any of the peptidase families previously reported and should, therefore, be regarded as a prototype enzyme that defines a new family of dipeptidylpeptidases.

The invention further includes a polypeptide, preferably a dipeptidylpeptidase, that shares a significant level of primary structure (referred to as "percent identity") with SEQ ID NO:2. The level of identity is determined

by aligning the two amino acid sequences (i.e., the amino acid sequence of the polypeptide and the sequence SEQ ID NO:2) such that the residues that make up the putative active site sequence (e.g., about amino acid 644 to about 653, preferably about amino acid 644 to about 658) are in register, then further aligned to maximize the number of amino acids that they have in common along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to place the residues that make up the putative active site sequence (e.g., about amino acid 644 to about 653, preferably about amino acid 644 to about 658) in register and to maximize the number of shared amino acids, although the amino acids in each sequence must nonetheless remain in their proper order. Preferably, two amino acid sequences are compared using the blastp program of the BLAST search algorithm, which is described by Altshul et al., (*Nucl. Acids Res.*, 25, 3389-3402 (1997)), and available at the National Center for Biotechnology Information (e.g., www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html or www.ncbi.nlm.nih.gov/BLAST/). Preferably, the default values for all BLAST search parameters are used. In the comparison of two amino acid sequences using the BLAST search algorithm, structural similarity is referred to as "identities." Preferably, a dipeptidylpeptidase has, in increasing order of preference, at least about 40% identity, at least about 50% identity, at least about 60% identity, at least about 70% identity, at least about 80% identity, and most preferably at least about 90% identity with SEQ ID NO:2. Preferably, about amino acid 543 to about 699 of SEQ ID NO:2 are used, more preferably about amino acid 71 to about 712 of SEQ ID NO:2 are used. Preferably the invention includes an isolated polypeptide including an amino acid sequence having a percentage amino acid identity of greater than about 40% with SEQ ID NO:2.

In general, the amidolytic activity of the polypeptides of the invention, preferably dipeptidylpeptidases, can be measured by assay of the cleavage of a target polypeptide in the presence of dipeptidylpeptidase and a buffer. Preferably, the ratio of dipeptidylpeptidase to target polypeptide is at least about 1:1; more preferably at least about 1:100; even more preferably at least about 1:1,000; and most preferably at least about 1:10,000. Preferably, the ratio of

dipeptidylpeptidase to target polypeptide is at most about 1:10,000,000; more preferably at most about 1:1,000,000; and most preferably at most about 1:100,000. Buffers in which a dipeptidylpeptidase is active are suitable for the assay. Preferably, the buffer is at most about 200 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), more preferably at most about 50 mM HEPES, and most preferably at most about 20 mM HEPES. Preferably, the pH of the buffer is at least about pH 6.0. Preferably, the pH of the buffer is at most about pH 8.0 and more preferably at most about pH 7.5. Preferably, the temperature of the assay is about 37°C. The assay can be carried out for at least about 1 minute to at most about 24 hours. Preferably, the amidolytic activity of the dipeptidylpeptidases are measured at a dipeptidylpeptidase:target polypeptide ratio of at least about 1:100 and at most about 1:1,000,000 in about 200 mM HEPES, about pH 7.5 at about 37°C for at least about 3 hours. In general, the time of the assay can vary depending on the substrate and enzyme:substrate ratio. Typically, target polypeptides are stable under these conditions, and typically it is difficult to detect background levels of hydrolysis in the absence of a dipeptidylpeptidase. Preferably, the assay is allowed to continue until at least about 1% of the target polypeptide is hydrolyzed.

Dipeptidylpeptidases of the present invention are preferably inhibited by compounds including, for example, 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (PEFABLOCK); diisopropylfluorophosphate (DFP); phenylmethanesulfonyl fluoride (PMSF); 3,4-dichloisocoumarin; and combinations thereof. The peptidases of the present invention are preferably not inhibited by a compounds including, for example, specific inhibitors of metallo peptidases, cysteine peptidases, and aspartic peptidases.

An active analog, active fragment, or active modification of a polypeptide including the amino acid sequence SEQ ID NO:2 is one that has amidolytic activity by hydrolysis of the target polypeptide described above. Active analogs of a polypeptide including the amino acid sequence SEQ ID NO:2 include dipeptidylpeptidases having amino acid substitutions that do not eliminate hydrolysis of the target polypeptide at the peptide bond between the second and third amino acids. Substitutes for an amino acid may be selected

from other members of the class to which the amino acid belongs. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, and tryptophan. Polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Examples of preferred conservative substitutions include Lys for Arg and *vice versa* to maintain a positive charge; Glu for Asp and *vice versa* to maintain a negative charge; Ser for Thr so that a free -OH is maintained; and Gln for Asn to maintain a free NH_2.

Active fragments of a dipeptidylpeptidase of the invention include dipeptidylpeptidases containing deletions or additions of one or more contiguous or noncontiguous amino acids such that the resulting polypeptide will hydrolyze the target polypeptide at the bond between the second and third amino acids. Modified dipeptidylpeptidases include dipeptidylpeptidases that are chemically and enzymatically derivatized at one or more constituent amino acid, including side chain modifications, backbone modifications, and N- and C- terminal modifications including acetylation, hydroxylation, methylation, amidation, and the attachment of carbohydrate or lipid moieties, cofactors, and the like. Modified dipeptidylpeptidases will hydrolyze the target polypeptide at the peptide bond between the second and third amino acids.

Preferably, a dipeptidylpeptidase includes the sequence TGGNSGSPV (SEQ ID NO:26), and more preferably includes the consensus sequence for the active-site serine residue of serine type proteases, TGGNSGSPVF (SEQ ID NO:25), where T is Threonine, G is glycine, N is Asparagine, P is Proline, V is valine, F is Phenylalanine, and S is serine, with the putative active site serine being underlined. The active site serine can be identified by, for instance, labeling with diisopropylfluorophosphate as described herein. Preferably, the catalytic domain of the dipeptidylpeptidases of the invention begins at about residue 543 of SEQ ID NO:2 and includes the remaining 169 amino acids, more preferably begins at about residue 540 of SEQ ID NO:2 and includes the

remaining 172 amino acids, and most preferably begins at about residue 522 of SEQ ID NO:2 and includes the remaining carboxy-terminal amino acids.

Dipeptidylpeptidases can be obtained by several methods. Isolation of a dipeptidylpeptidase present on the surface of a cell producing the peptidase typically requires lysis of the cell followed by purification methods that are well known in the art. Alternatively, cells can be treated with a detergent, for instance Triton X-100, to remove the peptidase from the cell surface. The following are nonlimiting examples of suitable protein purification procedures: fractionation on immunoaffinity, ion-exchange, hydroxyapatite, Phenyl-Sepharose HP, MonoQ HR 5/5, or MonoP columns; ethanol precipitation; reverse phase HPLC; chromatography on silica or on an ion-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75. Preferably, isolation of a dipeptidylpeptidase from *P. gingivalis* is accomplished using a combination of hydroxyapatite, Phenyl-Sepharose HP, MonoS HR 5/5 and MonoP column chromatography steps as described herein.

Dipeptidylpeptidases can also be isolated from organisms other than *P. gingivalis*. Other organisms can express a dipeptidylpeptidase that is encoded by a coding region having similarity to the coding region encoding SEQ ID NO:2. A "coding region," a "coding sequence," or an "open reading frame" (ORF) is a linear form of nucleotides that encodes a polypeptide, usually via mRNA, when placed under the control of appropriate regulatory sequences. The boundaries of a coding region are generally determined by a translation start codon at its 5' end and a translation stop codon at its 3' end. "Regulatory region" refers to a nucleic acid that regulates expression of a coding region to which a regulatory region is operably linked. Non limiting examples of regulatory regions include promoters, transcription initiation sites, translation start sites, translation stop sites, and terminators. "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A regulatory element is "operably linked" to a coding region when it is joined in such a way that expression of the coding region is achieved under conditions compatible with the regulatory region. Alternatively,

other organisms can express a dipeptidylpeptidase from a recombinant coding region encoding the peptidase. The identification of similar coding regions in other organisms can be accomplished as described herein. A dipeptidylpeptidase can be isolated using purification methods that are well known in the art. Alternatively, the peptidase can be chemically synthesized using methods that are well known in the art including, for instance, solid phase synthesis. Examples of, for instance, coding and regulatory regions are described herein.

The expression of a dipeptidylpeptidase by an organism other than *P. gingivalis* can be detected using specific substrates of the general formula Xaa-Xaa-LG, wherein Xaa represents any natural amino acid and LG is a leaving group. The leaving group can be a chromogenic or fluorogenic group known to the art. The expression of a dipeptidylpeptidase by an organism and subsequent cleavage of a specific substrate results in a free amino acid or a free leaving group, each of which can be assayed using techniques known to those of skill in the art. Other methods can be based on immunogenic properties of DPP-7, for instance immunoassays and histochemistry, the detection of mRNA, and PCR related methods, all of which are known to one of skill in the art.

In one aspect, the present invention is directed to a nucleic acid encoding a polypeptide, particularly a dipeptidylpeptidase, active analog, active fragment, or active modification thereof. The nucleic acid can have a nucleotide sequence as shown in SEQ ID NO:1. Alternatively, nucleic acids of the invention include those whose complement hybridize to SEQ ID NO:1 under standard hybridization conditions as described herein. During hybridization the entire nucleotide sequence of the complement can hybridize with SEQ ID NO:1. Preferably, at least about 20 nucleotides of the complement hybridize with SEQ ID NO:1, more preferably at least about 50 nucleotides, most preferably at least about 100 nucleotides.

The identification of similar coding regions in other organisms can be accomplished by screening individual wild-type microorganisms for the presence of nucleotide sequences that are similar to the coding region of DPP-7, which is shown in SEQ ID NO:1. Screening methods include, for instance, hybridization of a detectably labeled probe with a nucleic acid.

Standard hybridizing conditions are a modification of the conditions used by Church et al. ((1984) *Proc. Natl. Acad. Sci. USA* 81, 1991): 0.5 M phosphate buffer, pH 7.2, 7% SDS, 10 mM EDTA, at 68°C, and three washes, each for 20 minutes in 2x SSC (preferably 0.1 SSC), 0.1% SDS, at 65°C. Preferably, a probe will hybridize to the nucleotide sequence set forth in SEQ ID NO:1 under standard hybridizing conditions. Generally the probe does not have to be complementary to all the nucleotides of the nucleic acid as long as there is hybridization under the above-stated conditions.

"Complement" and "complementary" refer to the ability of two single stranded nucleic acids to base pair with each other, where an adenine on one nucleic acid will base pair to a thymine on a second nucleic acid and a cytosine on one nucleic acid will base pair to a guanine on a second nucleic acid. Two nucleic acids are complementary to each other when a nucleotide sequence in one nucleic acid can base pair with a nucleotide sequence in a second nucleic acid. For instance, 5'-ATGC and 5'-GCAT are complementary. The term complement and complementary also encompasses two nucleic acids where one nucleic acid contains at least one nucleotide that will not base pair to at least one nucleotide present on a second nucleic acid. For instance the third nucleotide of each of the two nucleic acids 5'-ATTGC and 5'-GCTAT will not base pair, but these two nucleic acids are complementary as defined herein. Typically two nucleic acids are complementary if they hybridize under the standard conditions referred to herein.

Preferred probes are nucleic acids complementary to a coding region or another nucleotide sequence that encodes a dipeptidylpeptidase. For instance, a probe can include a consecutive series of nucleotides complementary to a portion of SEQ ID NO:1. Preferably a probe is at least about 18 bases, more preferably at least about 21 bases, and most preferably at least about 24 bases in length. One of skill in the art could select useful probes as desired. Methods of detectably labeling a probe are well known to the art.

The nucleic acid that is identified by the probe is further analyzed to determine if it encodes a polypeptide with amidolytic activity of the peptide bond between the second and third amino acids from the N-terminus on a target polypeptide of the general formula H-Xaa-Yaa-Xaa, wherein Xaa is a natural or modified amino acid and Yaa is an amino acid including an aliphatic or an aromatic residue as a substituent on the α-carbon atom. Another method for screening individual microorganisms for the presence of nucleotide sequences that are similar to the coding regions of the present invention is the polymerase chain reaction (PCR).

Individual wild-type microorganisms containing nucleic acids encoding a dipeptidylpeptidase can also be identified using antibody. Preferably the antibody is directed to DPP-7. The production of antibodies to a particular polypeptide is known to a person of skill in the art, and is further detailed herein.

The use of hybridization of a probe to a coding region present in individual wild-type microorganisms can be used as a method to identify a coding region identical or similar to a coding region present in SEQ ID NO:1. The coding region can then be isolated and ligated into a vector as described below.

The present invention is also directed to coding regions sharing a significant level of primary structure with the coding region present at SEQ ID NO:1. The level of identity is determined by aligning the two nucleotide sequences (i.e., the nucleotide sequence of the polynucleotide and the sequence SEQ ID NO:1) such that the residues that encode the putative active site of the encoded protein (e.g., about nucleotide 1929 to about 1974) are in register, then further aligned to maximize the number of nucleotides that they have in common along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to place the residues that encode the putative active site of the encoded protein (e.g., about nucleotide 1929 to about 1974) in register and to maximize the number of shared nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. Preferably, two nucleotide sequences are compared using the blastn program of the BLAST search algorithm, which is described by Altshul et al., (*Nucl. Acids*

Res., 25, 3389-3402 (1997)), and available at the National Center for Biotechnology Information (e.g., www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html or www.ncbi.nlm.nih.gov/BLAST/). Preferably, the default values for all BLAST search parameters are used. In the comparison of two nucleotide sequences using the BLAST search algorithm, structural similarity is referred to as "identities." Preferably, two nucleotide acid sequences have, in increasing order of preference, preferably at least about 70%, at least about 80%, at least about 90%, at least about 95%, and most preferably at least about 95% identity.

As mentioned above, a nucleic acid of the invention can be inserted in a vector. Construction of vectors containing a nucleic acid of the invention employs standard ligation techniques known in the art. See, e.g., Sambrook et al, *Molecular Cloning: A Laboratory Manual.*, Cold Spring Harbor Laboratory Press (1989) or Ausubel, R.M., ed. *Current Protocols in Molecular Biology* (1994). A vector can provide for further cloning (amplification of the nucleic acid), i.e., a cloning vector, or for expression of the polypeptide encoded by the coding region, i.e., an expression vector. The term vector includes, but is not limited to, plasmid vectors, viral vectors, cosmid vectors, or artificial chromosome vectors. Typically, a vector is capable of replication in a bacterial host, for instance *E. coli*. Preferably the vector is a plasmid.

Selection of a vector depends upon a variety of desired characteristics in the resulting construct, such as a selection marker, vector replication rate, and the like. Suitable plasmids for expression in *E. coli*, for example, include pUC(X), pKK223-3, pKK233-2, pTrc99A, and pET-(X) wherein (X) denotes a vector family in which numerous constructs are available. pUC(X) vectors can be obtained from Pharmacia Biotech (Piscataway, NH) or Sigma Chemical Co. (St. Louis, MO). pKK223-3, pKK233-2 and pTrc99A can be obtained from Pharmacia Biotech. pET-(X) vectors can be obtained from Promega (Madison, WI) Stratagene (La Jolla, CA) and Novagen (Madison, WI). To facilitate replication inside a host cell, the vector preferably includes an origin of replication (known as an "ori") or replicon. For example, ColE1 and P15A replicons are commonly used in plasmids that are to be propagated in *E. coli*.

An expression vector optionally includes regulatory regions operably linked to the coding region. The invention is not limited by the use of any particular promoter, and a wide variety are known. Promoters act as regulatory signals that bind RNA polymerase in a cell to initiate transcription of a downstream (3' direction) coding region. The promoter used in the invention can be a constitutive or an inducible promoter. It can be, but need not be, heterologous with respect to the host cell. Preferred promoters for bacterial transformation include *lac*, *lac*UV5, *tac*, *trc*, T7, SP6 and *ara*.

An expression vector can optionally include a Shine Dalgarno site (e.g., a ribosome binding site), and a start site (e.g., the codon ATG) to initiate translation of the transcribed message to produce the enzyme. It can also include a termination sequence to end translation. A termination sequence is typically a codon for which there exists no corresponding aminoacetyl-tRNA, thus ending polypeptide synthesis. The nucleic acid used to transform the host cell can optionally further include a transcription termination sequence. The rrnB terminators, which is a stretch of DNA that contains two terminators, T1 and T2, is an often used terminator that is incorporated into bacterial expression systems (J. Brosius et al., (1981) *J. Mol. Biol.* 148 107-127).

The nucleic acid used to transform the host cell optionally includes one or more marker sequences, which typically encode a polypeptide that inactivates or otherwise detects or is detected by a compound in the growth medium. For example, the inclusion of a marker sequence can render the transformed cell resistant to an antibiotic, or it can confer compound-specific metabolism on the transformed cell. Examples of a marker sequence are sequences that confer resistance to kanamycin, ampicillin, chloramphenicol, and tetracycline.

Antibodies to a polypeptide including the sequence SEQ ID NO:2 can be produced. Alternatively, antibodies to an antigenic analog, antigenic fragment, or antigenic modification of a polypeptide including the sequence SEQ ID NO:2 can be made. An antigenic analog, antigenic fragment, or antigenic modification of a polypeptide including the amino acid sequence SEQ ID NO:2 is one that generates an immune response in an animal. Preferably, an antigenic analog, antigenic fragment, or antigenic modification has amidolytic activity. Antigenic

analogs of a polypeptide including the amino acid sequence SEQ ID NO:2 include dipeptidylpeptidases having amino acid substitutions that do not eliminate peptide antigenicity in an animal. Substitutes for an amino acid may be selected from other members of the class to which the amino acid belongs, as described herein. Fragments of a dipeptidylpeptidase of the invention include dipeptidylpeptidases containing deletions or additions of one or more contiguous or noncontiguous amino acids such that the resulting polypeptide will generate an immune response in an animal. Modified dipeptidylpeptidases include dipeptidylpeptidases that are chemically and enzymatically derivatized at one or more constituent amino acids, including side chain modifications, backbone modifications, and N- and C- terminal modifications including acetylation, hydroxylation, methylation, amidation, and the attachment of carbohydrate or lipid moieties, cofactors, and the like.

Accordingly, an aspect of the invention is an immunogenic composition including an isolated dipeptidylpeptidase, an antigenic analog, antigenic fragment, or antigenic modification thereof. The dipeptidylpeptidase preferably has amidolytic activity for cleavage of the target polypeptide described herein.

The immunogenic composition can further include excipients or diluents that are pharmaceutically acceptable as carriers and compatible with the immunogenic composition. The term "pharmaceutically acceptable carrier" refers to a carrier(s) that is "acceptable" in the sense of being compatible with the other ingredients of a composition and not deleterious to the recipient thereof. Suitable excipients include, for example, water, saline, dextrose, glycerol, ethanol, and combinations thereof. In addition, if desired, the immunogenic composition may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the immune-stimulating composition.

The immunogenic composition can be used in a method for protecting an animal from a disease caused by *P. gingivalis*. This method includes administering the immunogenic composition and eliciting antibodies to a dipeptidylpeptidase, antigenic analog, antigenic fragment, or antigenic modification. The diseases that can be treated in this manner include

periodontal diseases, which include gingivitis and periodontitis. Clinical hallmarks of periodontitis include loss of tooth attachment and periodontal pocket formation.

Alternatively and preferably, periodontal diseases can be treated by the use of inhibitors of a dipeptidylpeptidase. An inhibitor of a dipeptidylpeptidase can be present in a composition that preferably contains a pharmaceutically acceptable carrier. For instance, inhibitors can be applied systemically, subgingivally (e.g., subgingival irrigation), and/or by controlled release delivery directly into the periodontal pocket using methods well known in the art (see, e.g., Kornman, *J. Periodontol.* 64:782-91 (1993). Preferably, an inhibitor is applied subgingivally or by controlled release delivery.

The dipeptidylpeptidases, active analogs, active fragments, and active modifications thereof can be used in a method of reducing growth of bacteria *in vitro* or *in vivo*. Preferably, the bacteria is a periodontal pathogen, i.e, a bacterial pathogen that causes periodontal disease, more preferably the bacteria is *P. gingivalis*. The inability of asaccharolytic *P. gingivalis* to utilize free amino acids makes the bacterium entirely dependant on an external peptide supply. The action of the polypeptides of the invention may be required for bacterial growth, and inhibition of the polypeptides of the invention may inhibit the *in vivo* growth of organisms, including *P. gingivalis*. The method includes decreasing the amount of dipeptides (e.g., the result of cleavage of the target polypepitde by a dipeptidylpeptidase) and the amount of free amino acids that result from further cleavage of the dipeptides present by inhibiting a dipeptidylpeptidase, active analog, active fragment, or active modification thereof, such that the amount of dipeptides generated by the polypeptides is decreased. The amount of dipeptides is decreased relative to the amount of dipeptides present in the absence of the inhibitor. Preferably, the amount of dipeptides generated is decreased by an inhibitor, a monoclonal antibody that inhibits the dipeptidylpeptidase, or polyclonal antibodies that inhibit the dipeptidylpeptidase, more preferably, the amount of dipeptides generated is decreased by an inhibitor. Preferably, an inhibitor acts to inhibit a polypeptide of the invention, preferably a dipeptidylpeptidase, by blocking the active site of the

polypeptide. The polypeptide can be present on the surface of the bacteria or secreted into the environment, preferably the polypeptide is present in the surface of the bacteria.

The present invention is also directed to a method of developing an inhibitor of a dipeptidylpeptidase, active analog, active fragment, or active modification thereof, preferably a dipeptidylpeptidase. The method includes identifying a compound that inhibits the amidolytic activity of the dipeptidylpeptidase. Such compounds include, for example, polypeptides, organic compounds, inorganic compounds, metals, non-ribosomal polypeptides, polyketides, and peptidomimetics. The identification of compounds can be accomplished by, for instance, incubating the dipeptidylpeptidase with a candidate compound under conditions that promote amidolytic activity of the dipeptidylpeptidase and determining if the amidolytic activity of the dipeptidylpeptidase is decreased relative to the amidolytic activity in the absence of the compound. The amidolytic activity can be measured by cleavage of the peptide bond between the second and third amino acids of the target polypeptide as described herein. One method of developing an inhibitor includes using the target polypeptide and replacing the Xaa residues with modified amino acids. It is expected that some modified amino acids will cause the target polypeptide to act as an inhibitor.

The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the invention as set forth herein.

EXAMPLES

Materials

Diisopropylfluorophosphate (DFP), leupeptin and 3,4-dichloroisocoumarin, were purchased from Calbiochem (La Jolla, CA). Other peptides used in this study were synthesized at the Molecular Genetic Instrumental Facility (University of Georgia, Athens, GA) using Fmoc protocol

with an advanced ChemTech MPS350 automated synthesizer. The peptides H-Xaa-Xaa-pNA and Z-Xaa-Xaa-pNA, where pNA is p- Nitroanilide; Z is benzyloxycarbonyl; and H is hydrogen and denotes an unblocked amino-terminal group were obtained from Bachem (King of Prussia, PA).

Methods

EXAMPLE 1

Source and Cultivation of Bacteria - P. gingivalis DPP-7 was purified from strain HG66, a kind gift of Dr. Roland Arnold (University of North Carolina, Chapel Hill, NC). The cells were grown as described previously (Chen et al., *J. Biol. Chem.*, 267:18896-901 (1992)).

Protein Determination - Protein concentration was determined with the BCA reagent kit (Sigma), using bovine serum albumin as a standard.

Localization of Dipeptidylpeptidase Activity - The localization of active enzyme was checked in bacterial cells that had been subjected to a previously described fractionation procedure (Banbula et al., *Infect. Immun.*, 68:1176-82 (2000)). All fractions, as well as the full culture, culture medium, and full culture after sonication, were assayed for amidolytic activity against H-Ala-Phe-pNA.

Enzyme Purification - All purification steps were performed at 4°C except for FPLC separations, which were carried out at room temperature. The cells were collected by centrifugation (6,000 x g, 30 minutes) and resuspended in 50 mM potassium phosphate buffer, pH 7.4. The outer membrane proteins were solubilized with 0.05% Triton X-100. After 2 hours of gentle stirring, unbroken cells were removed by centrifugation (28,000-x g, 60 minutes). Proteins from the supernatant were precipitated with cold acetone (60% final concentration), collected by centrifugation, and redissolved in 50 mM potassium phosphate buffer, pH 7.0. After extensive dialysis against the same buffer the sample was loaded onto a hydroxyapatite column (BioRad) previously equilibrated with 20 mM potassium phosphate, pH 7.0, at a flow rate of 20 ml/hour. The column was then washed until the A_{280} fell to zero. Bound proteins were eluted with a potassium phosphate gradient (20-300 mM) and fractions (7 ml) were analyzed

for amidolytic activity against H-Ala-Phe-pNA. The active fractions were saturated with 1 M ammonium sulfate and loaded onto a Phenyl-Sepharose HP column (Pharmacia) equilibrated with 50 mM potassium phosphate, pH 7.0, containing 1M ammonium sulfate. The column was washed with two volumes of the equilibration buffer, followed by a wash with buffer containing 0.4 M ammonium sulfate, and developed with a descending gradient of ammonium sulfate from 0.4 to 0 M. Active fractions were pooled, extensively dialyzed against 20 mM MES, pH 6.6, and applied onto a MonoS HR 5/5 FPLC (Pharmacia) column equilibrated with the same buffer. Bound proteins were eluted with a 0-300 mM NaCl gradient. This allowed us to obtain a homogenous preparation of active proteinase.

Electrophoretic Techniques - The SDS-PAGE system of Schagger and von Jagow (Schagger et al., *Anal. Biochem.*, 166:368-79 (1987)), was used to monitor enzyme purification and estimate the enzyme molecular mass. For amino-terminal sequence analysis, proteins resolved in SDS-PAGE were electroblotted onto polyvinylidene difluoride membranes using 10 mM CAPS, pH 11, 10% methanol (Matsudaira et al., *J. Biol. Chem.*, 262:10035-38 (1987)). After staining with Coomasie Blue G250 the blot was air dried, and protein bands cut out and subjected to amino-terminal sequence analysis with an Applied Biosystems 491 Protein Sequencer using the program designed by the manufacturers.

Kinetic Analysis - Routinely, the dipeptidylpeptidase amidolytic activity was measured with H-Ala-Phe-pNA (1mM) in 0.2 M HEPES, pH 7.8 at 37°C. The reaction was followed for specific time intervals in a thermostated ELISA reader (SpectraMax, Applied Biosystem) and the release of p-nitroaniline was monitored at 405 nm. Other p-nitroanilide substrates were used in the same manner. For inhibition studies, the enzyme was first preincubated with an inhibitor for 15 minutes at 37°C, substrate added, and residual activity recorded. The initial steady-state velocity (v_0) was determined by continuous assay for the range of substrate concentrations (100 nM to 1 mM). K_m and V_{max} were determined by hyperbolic regression of the kinetic data using the software

package Hyper Version 1.02 obtained from Dr. J. S. Easterby (University of Liverpool, UK).

Enzyme Fragmentation - The purified dipeptidylpeptidase was subjected to in-gel tryptic digestion (Rosenfeld et al., *Anal. Biochem.*, 203:173-79 (1992)). Peptides were extracted and separated by microbore reverse-phase HPLC. Fractions absorbing at 210 nm were manually collected, and their masses were determined by reflectron MALDI-TOF mass spectrometry using a Bruker Daltonics ProFlex instrument as described previously (Pohl et al., *Lett. Peptide Sci.*, 1:291-97 (1995)). Selected peptides were subjected to Edman degradation in a model Procise-cLS sequencer (PE Biosystems, CA).

Identification of the DPP-7 Coding Sequence - An unfinished *P. gingivalis* W83 genome database, available from the Institute for Genomic Research, was searched for the presence of nucleotide sequences corresponding to the amino-terminal and the internal DPP-7 amino acid sequences using the TBLASTN algorithm (Altschul et al., *Nucleic Acids Res.*, 25:3389-402 (1997)). An identified contig gnl | TIGR | *P. gingivalis*_1208 was retrieved from the Institute for Genomic Research database. The position of the DPP-7 coding sequence was localized using the National Center for Biotechnology Information (NCBI) open reading frame (ORF) finder and the amino acid sequence, obtained by conceptual translation of the entire ORF, was further used for homology screening by use of the NCBI BLAST search tool.

Enzyme Specificity - The determination of substrate specificity was based on the separation of the products of peptide hydrolysis by reverse-phase chromatography. Peptides were first incubated with 1 microgram of DPP-7 at an enzyme: substrate molar ratio of 1:100 for 3 hours or 24 hours in 50 microliters of 200 mM HEPES, 100 mM NaCl pH 8.0, at 37°C, and the reaction stopped by acidification with trifluoroacetic acid. The samples were then subjected to reverse-phase high pressure liquid chromatography using a Supelcosil LC 18 column (Supelco) with an acetonitrile gradient 0-60% in 0.075% trifluoroacetic acid in 50 minutes. Each peak, detected at 210 nm, was collected, lyophilized, re-dissolved in 50% (v/v) methanol, 0.1% acetic acid and subjected to analysis by mass spectrometry.

EXAMPLE 2

A 76 kDa dipeptidylpeptidase associated with *P. gingivalis* membranes was solubilized by mild detergent treatment. This procedure released more than 90% of the amidolytic activity against H-Ala-Phe-pNA into the medium. After acetone precipitation and subsequent chromatography steps including the use of hydroxyapatite, Phenyl-Sepharose and MonoS columns (Fig. 1) a pure enzyme preparation was obtained. The homogeneity of the preparation and molecular mass of the protein were checked both by SDS PAGE (Fig. 2) and gel filtration on a TSK G3000 SW column.

EXAMPLE 3

Inhibition Profile - Based on the inhibition studies (Table I), DPP-7 was classified as a serine protease. DPP-7 was inactivated by diisopropylfluorophosphate, PEFABLOCK and 3,4-dichloisocoumarin, but not by typical cysteine class inhibitors such as E-64 or iododoacetic acid. Metal chelators including EDTA and 1,10-orthophenanthroline, as well as reducing agents did not influence its activity. The enzyme was not sensitive to inactivation by either detergents (0.5% SDS, 1% Triton X-100) or heavy metal ions including Zn^{2+}, Co^{2+} and Ni^{2+}. Human plasma inhibitors, such as α_1-proteinase inhibitor, α_1-antichymotrypsin, and α_2-macroglobulin, did not effect enzyme activity nor were they cleaved by DPP-7.

TABLE I: Effect of different compounds on P. gingivalis DPP-7 activity.

Inhibitor	Concentration	% of residual activity
Diisopropylfluorophosphate	10 mM	34
PEFABLOCK	4 mg/ml	1
3,4-dichloroisocoumarin	2 mM	0
E-64	1 micromolar	96
Iodoacetic acid	0.1 mM	102
EDTA	10 mM	90
1,10-orthophenanthroline	1 mM	98

Leupeptin	0.1 mM	107
Aprotinin	0.5 mg/ml	128
Pepstatin	0.5 mg/ml	127
Cysteine	10 mM	90
Gly-Ala	100 mM	102
Arg-Phe	100 mM	69
Ala-Gly	100 mM	96
Arg-Gly	10 mM	84
Lys-Gly	10 mM	96
Ni^{++}	1 mM	95
Zn^{++}	1 mM	95
Co^{++}	1 mM	116
SDS	0.5%	65
SDS	1%	0
Triton X-100	0.1%	144
Triton X-100	0.5%	103
Triton X-100	1%	94

EXAMPLE 4

pH Optimum and Stability - Purified DPP-7 was active against H-Ala-Phe-pNA over a broad pH range, from neutral to basic pH (6.5-9.0) (Fig. 3). This activity also changed with the ionic strength of the buffer, reaching 200% at 0.5 M NaCl concentration in 100 mM HEPES, pH 8.0. DPP-7 was stable in 0.2 M HEPES, pH 8.0, for one week at 4°C. The protease showed no appreciable loss of activity when kept frozen at -80°C for one month. After 3 hours incubation at either room temperature or 37°C, activity was reduced to 62% and 20%, respectively. The optimum temperature for the hydrolysis of H-Ala-Phe-pNA was determined to be 43°C.

EXAMPLE 5

Substrate Specificity - Among several chromogenic substrates tested, only those with an aliphatic or an aromatic side chain residues in the second, penultimate position were rapidly hydrolyzed by DPP-7 (Table 2).

TABLE 2: Kinetic analysis for paranitroanalides cleavage by DPP-7.

Substrate	Km [mM]	Vmax
H-Ala-Ala-pNA	0.313	129.65
H-Ala-Phe-pNA	0.441	170.06
H-Gly-Phe-pNA	0.256	54.54

Several other substrates including H-Ala-Pro-pNA, H-Ala-pNA, H-Gly-pNA, H-Ile-pNA, H-Leu-pNA, H-Lys-pNA, H-Phe-pNA, H-Gly-Arg-pNA, H-Gly-Glu-pNA, H-Gly-Lys-pNA, H-Ala-Gly –pNA, H-Gly-Gly-pNA, H-Ala-Ala-Phe-pNA,H-Ala-Gly-Arg-pNA, H-Leu-Thr-Arg-pNA, H-Ala-Phe-Pro-pNA, Nα-benzoyl-DL-arginine-pNA, N-met-Ala-Pro-Val-pNA, N-suc-Ala-Ala-pNA, N-suc-Ala-Ala-Pro-Glu-pNA, N-suc-Ala-Ala-Pro-Leu-pNA, N-suc-Ala-Ala-Val-Ala-pNA, Z-Ala-Ala-pNA, Z-Lys-pNA, Z-Arg-pNA, Z-Glu-Glu-pNA, Z-Leu-Leu-Glu-pNA, Z-Lys-Arg-pNA, Z-Phe-Arg-pNA, Z-Phe-Val-Arg-pNA, Z-Tyr-Lys-Arg-pNA were tested, but none of these was hydrolysed by DPP-7.

To further confirm specificity, several synthetic peptides were also tested as substrates for this enzyme. Again, only those polypeptides having an amino acid with an aliphatic or an aromatic residue as a substituent on the α-carbon atom of the second amino acid from the N-terminal end of the polypeptide were cleaved (Table 3), with glycine, proline, or charged amino acids not being acceptable as the second amino acid from the N-terminal end of the polypeptide. The protease did not show any endopeptidase activity on gelatin, insulin β chain, carboxymethylated lysosyme, azocazein or type I collagen. Purified DPP-7 was

devoid of any aminopeptidase activity and did not cleave model substrates with blocked amino-termini.

TABLE 3: Specificity of P. gingivalis DPP-7 on synthetic peptides.

Peptides cleaved	Peptides not cleaved
Trp-Ala-↓-Gly-Gly-Asp-Ala-Ser-Gly-Glu (SEQ ID NO:10)	Trp-His-Trp-Leu-Glu-Leu-Lys-Pro-Gly-Glu-Pro-Met-Tyr (SEQ ID NO:14)
Ile-Ala-↓-Arg-Arg-His-Pro-Tyr-Phe-Leu (SEQ ID NO:11)	Ser-Pro-Tyr-Ser-Ser-Glu-Thr-Thr (SEQ ID NO:15)
Lys-Ile-↓-Ala-Gly-Tyr-His-Leu-Glu-Leu (SEQ ID NO:12)	Ala-Pro-Val-Arg-Ser-Leu (SEQ ID NO:16)
Phe-Leu-↓-Arg-Glu-Pro-Val-Ile-Phe-Leu (SEQ ID NO:13)	Gln-Lys-Gln-Met-Ser-Asp-Arg-Arg-Glu (SEQ ID NO:17)

An arrow indicates cleavage site

EXAMPLE 6

DPP-7 Sequence Analysis - Purified DPP-7 was resolved on SDS-PAGE and electroblotted onto a PVDF membrane. It had an amino-terminal sequence ADKGMMWLLNELNQENLDRMRELGFT (SEQ ID NO:18). After proteolytic in-gel digestion of the enzyme additional internal sequences were obtained, including: DNKPYK (SEQ ID NO:19), EMTYL (SEQ ID NO:20), FAQFAN (SEQ ID NO:21), VLPAML (SEQ ID NO:22), SVVPY (SEQ ID NO:23), LFFAGL (SEQ ID NO:24). All of this sequence data allowed us to identify the *P. gingivalis* genomic contig gln | TIGR | *P. gingivalis*_ in the Unfinished Microbial Genomes database, TIGR. An ORF corresponding to the DPP-7 amino acid sequence (SEQ ID NO:1) was found, as indicated by the fact, that all sequences of the DPP-7 derived peptides obtained by the enzyme polypeptide fragmentation by trypsin were present in the protein primary structure inferred from the nucleotide sequence of the ORF as shown in Fig. 4. Including a signal peptide (residues 1-24), the entire ORF corresponds to a 712 amino acid polypeptide (see Fig. 4). Interestingly, the DPP-7 ORF contains the

consensus sequence for the active-site serine residue of serine type proteases, TGGNSGSPVF (SEQ ID NO:25). As indicated in Fig. 5 the DPP-7 carboxy-terminus (SEQ ID NO:3) exhibits high degree of identity to that of the V8 serine protease (SEQ ID NO:4), particularly around the putative active site serine residue. This is surprising since the *P. gingivalis* DPP-7 is a dipeptidylpeptidase specific for substrates having an aliphatic or an aromatic residue as a substituent on the α-carbon atom of the second amino acid from the N-terminal end of the substrate, whereas *Staphyloccocus aureus* V8 endopeptidase is specific towards substrates including glutamic acid or aspartic acid as the second amino acid from the N-terminal end of the substrate. The similarity search performed using the NCBI TBLASTN tool against GenBank, EMBL, DDBJ and PDB databases showed no significant similarity of DPP-7 to any other known dipeptidylpeptidases, indicating that this enzyme could be regarded as a member of a new family of proteases. Additional searches against databases containing unfinished and finished microbial genomes allowed us to identify more coding sequences encoding similar proteases with consensus active site sequence TGGNSGSPV (Fig. 6; SEQ ID NO:26). A coding sequence of related protein has been found in *P. gingivalis* W83 unfinished portion of complete genome between positions 1360759 and 1362718. This putative proteinase reveals significant similarity to DPP-7 (267/691 identities). Another organism *Shewanella putrefaciens* possesses two related coding sequences (gnl |TIGR_24|sputre 6401 and gnl |TIGR_24|sputre 6410) while a plant pathogen *Xylella fastidiosa* contains one coding sequence encoding similar proteinase (gb|AE004008.1|). In addition, the computer assisted search for sequential motifs characteristic for transmembrane domains revealed the presence of two such putative regions within the amino-terminal sequence of DPP-7, with residues 7 to 24 and 62 to 78 most likely folded into hydrophobic α-helices responsible for membrane anchoring of this enzyme.

The complete disclosure of all patents, patent applications, and publications, and electronically available material (e.g., GenBank amino acid and nucleotide sequence submissions) cited herein are incorporated by reference.

The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

SEQUENCE LISTING FREE TEXT

SEQ ID NO:1	Coding sequence encoding *Porphyromonas gingivalis* DPP-7
SEQ ID NO:2	*Porphyromonas gingivalis* DPP-7
SEQ ID NO:3	C-terminal region of *Porphyromonas gingivalis* DPP-7
SEQ ID NO:4	C-terminal region of *Staphylococcus aureus V8* endopeptidase
SEQ ID NO:5	Coding sequence for *Shewanella putrefaciens* gnl | TIGR_24 | sputre 6401
SEQ ID NO:6	Coding sequence for *Shewanella putrefaciens* gnl | TIGR_24 | sputre 6410
SEQ ID NO:7	Coding sequence for *Xylella fastidiosa* gb | AE004008.1 |
SEQ ID NO:8	Coding sequence for *Porphyromonas gingivalis* gnl | TIGR | *P. gingivalis*_CPG.con
SEQ ID NO:9	Coding sequence for *Porphyromonas gingivalis* DPP-7 gnl | TIGR | *P. gingivalis*_CPG.con
SEQ ID NO:10-17	Synthetic peptides
SEQ ID NO:18	N-terminal region of *Porphyromonas gingivalis* DPP-7
SEQ ID NO:19-24	Internal sequences of *Porphyromonas gingivalis* DPP-7
SEQ ID NO:25-26	Consensus sequences for active sites for serine type proteases